John von Neumann Institute for Computing

Do Amyloidogenic Regions Intersect with Folding Nuclei of Native Structure?

O. V. Galzitskaya

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 219-222, 2008.

http://www.fz-juelich.de/nic-series/volume40

Do Amyloidogenic Regions Intersect with Folding Nuclei of Native Structure?

Oxana V. Galzitskaya

Institute of Protein Research, Russian Academy of Sciences,
Pushchino, Moscow Region, Russian Federation
E-mail: ogalzit@vega.protres.ru

A crucial event of protein folding is the formation of a folding nucleus, which is a structured part of the protein chain in the transition state. We demonstrate a correlation between locations of residues involved in the folding nuclei and locations of amyloidogenic regions. The average Φ-values are significantly greater inside amyloidogenic regions than outside them. We have found that fibril formation and normal folding involve many of the same key residues, giving an opportunity to outline the folding initiation site in protein chains.

1 Introduction

In spite of the fact that each protein has its own unique, native three-dimensional structure, some cases exist when there is another rather stable structure, called an amyloid fibril. Although native structures vary greatly from protein to protein, the structures of amyloid fibrils obtained from different proteins are fairly uniform. The formation of amyloid fibrils is a case of protein misfolding, in which a protein folds into a cross β-structure instead of folding into its native structure. In addition to proteins that form amyloid fibrils *in vivo* in various "amyloid diseases", there are many other proteins that are not implicated in amyloid diseases but form fibrils *in vitro*[1]. There is no sequence homology common to all such proteins or peptides.

Since polypeptide chains can fold into native structures or misfold into amyloid fibrils, there is a competition between the processes of folding and misfolding. During folding, a protein molecule has to overcome a free-energy barrier. The most unstable structure corresponds to the top of the barrier (i.e., to the transition state of the folding process)[2]. The folding nucleus is a structured part of the protein chain in the transition state. Since the folding nucleus is unstable, it is not easy to investigate it experimentally. A very laborious experimental method, which is called Φ-analysis, has been developed to determine the structure of folding nuclei[2].

The goal of this work is to compare amino acid residues which are crucial for folding and misfolding processes of the same proteins. As the experimental data on both folding nuclei and amyloidogenic regions in the same proteins are scarce, we compared experimentally found residues involved in folding nuclei with predicted residues involved in amyloidogenic regions and *vice versa*. We demonstrate that fibril formation and normal folding involve many of the same key residues. On average, Φ-values for amino acid residues in amyloidogenic regions are significantly greater than Φ-values for amino acid residues in non-amyloidogenic regions. This result allows us to search for some residues involved in the folding nucleus using only amino acid sequences.

2 Results and Discussion

2.1 Intersection of Predicted Residues Involved in Amyloidogenic Regions with Experimentally Found Residues Involved in Folding Nuclei

If amyloid fibril formation is a generic feature of proteins, some common properties of amino acid sequences possessing amyloidogenic propensities should be observed. Therefore, we can hypothesize that amyloidogenic regions often play a crucial role not only in the amyloid fibril formation but also in the process of "normal" folding of proteins into their native structure. We tested whether the experimentally found amino acid residues involved in folding nuclei intersect with theoretically predicted residues involved in amyloidogenic regions. The list of the experimentally found Φ-values as well as the corresponding mutations can be found at Ref. 3. Experimentally found Φ-values (larger 0.5) and predicted amyloidogenic regions can be found at Ref. 4.

We have compared predicted amyloidogenic regions with experimentally found residues involved in folding nuclei for those 20 proteins. The prediction of amyloidogenic regions was made by the previously described method which predicts amyloidogenic regions using only amino acid sequence[5,6]. For each amino acid residue, the method predicts the number of expected contacts and regions within which all residues have a large number of expected contacts are predicted as amyloidogenic ones. As it was demonstrated previously, this method is able to predict amyloidogenic regions[5].

The comparison of the degree of involvement into the folding nucleus (reflected in experimental Φ-values) of residues in the predicted amyloidogenic and non-amyloidogenic regions have demonstrated that there is a reliable difference. In the predicted amyloidogenic regions, the average over Φ-values is 0.41 ± 0.02 while in the predicted non-amyloidogenic regions, the average over Φ-values is 0.33 ± 0.01 (here and below, the shown error is the error of averaging which is calculated as $\frac{\sigma}{\sqrt{n}}$ where σ is the standard deviation of the distribution, and n is the number of points). Student's *t*-test gives the probability of 4×10^{-3}; thus, the above difference is statistically reliable.

Thus, comparison of experimentally known amino acid residues involved in the folding nuclei *vs.* predicted amyloidogenic fragments indicates that nucleation centers for folding and for misfolding often intersect.

2.2 Intersection of Experimentally Determined Amyloidogenic Regions with the Predicted Folding Nuclei

To investigate folding/unfolding behavior of amyloidogenic proteins, we have constructed a database of globular proteins with experimentally revealed amyloidogenic regions. From literature data, we selected those globular proteins in which the position of amyloidogenic regions is known from experimental data.6 The database now includes seven proteins: acylphosphatase, β2-microglobulin, gelsolin, transthyretin, lysozyme, myoglobin, human prion. We tested whether the theoretically found folding nuclei by our method[7] intersect with experimentally found amyloidogenic regions. It appears that 8 of 12 amyloidogenic regions are situated in folding nuclei where Φ-values are large. For several proteins, the regions with the largest Φ-values coincide with the amyloidogenic regions.

For amino acid residues in amyloidogenic regions, the average Φ-value is 0.58 ± 0.02 while amino acid residues in non-amyloidogenic regions have the average Φ-value that

is significantly smaller (0.43 ± 0.01). Thus, in amyloidogenic regions, an average amino acid residue has more than 50% of its contacts formed in the transition state (i.e., in the folding nucleus by our definition). The *p*-value obtained with Student's *t*-test (that is, the probability that the observed difference is accidental) is $2*10^{-11}$ that confirms that the difference between the average Φ-values of amino acid residues in amyloidogenic and in non-amyloidogenic regions is significant.

Thus, we have demonstrated that amyloidogenic regions are often predicted to be part of the folding nuclei in amyloidogenic proteins. Therefore, we can hypothesize that amyloidogenic regions often play a crucial role not only in amyloid fibril formation but also in the process of "normal" folding of amyloidogenic proteins into their native structure, since amyloidogenic regions compose part of the folding nucleus in these proteins.

Acknowledgments

We are grateful to S.O. Garbuzynskiy for assistance in some calculations. This work was supported by the program "Molecular and Cell Biology" of Russian Academy of Sciences, by the Russian Science Support Foundation, by the Russian Foundation for Basic Research, by the INTAS grant (05-1000004-7747), and by Howard Hughes Medical Institute (grant 55005607).

References

1. J. I. Guijarro, M. Sunde, J. A. Jones, I. D. Campbell and C. M. Dobson, *Amyloid fibril formation by an SH3 domain*, Proc. Natl. Acad. Sci. USA **95(8)**, 4224-4228, 1998.
2. J. T. Matouschek, Jr. Kellis, L. Serrano and A. R. Fersht, *Mapping the transition state and pathway of protein folding by protein engineering*, Nature **340(6229)**, 122-126, 1989.
3. `http://phys.protres.ru/resources/phi_values_amyloid.html`.
4. `http://phys.protres.ru/resources/phi_values.htm`.
5. O. V. Galzitskaya, S. O. Garbuzynskiy and M. Yu. Lobanov, *Prediction of amyloidogenic and disordered regions in protein chains*, PLoS Comput. Biol. **2(12)**, 4224-4228, 2006.
6. O. V. Galzitskaya, S. O. Garbuzynskiy and M. Yu. Lobanov,*Is it possible to predict amyloidogenic regions from sequence alone?*, Journal of Bioinformatics and Computational Biology **4(2)**, 373-388, 2006.
7. S. O. Garbuzynskiy, A. V. Finkelstein and O. V. Galzitskaya,*Outlining folding nuclei in globular proteins*, J Mol Biol **336(2)**, 509-525, 2004.